DOCUMENT RESUME

ED 458 263 TM 033 426

AUTHOR Wilson, Michael J.
TITLE The Assessment-Evaluation Cycle Model for Instruction.
PUB DATE 2001-00-00
NOTE 23p.
PUB TYPE Reports - Descriptive (141)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Decision Making; *Educational Assessment; Elementary Secondary Education; *Evaluation Methods; *Models; *Planning; Program Evaluation

ABSTRACT

A model is proposed that describes assessment-evaluation as practiced in instruction as a cycle. The model addresses consequences as an important culminating step in each assessment-evaluation cycle, so that the success of each cycle depends on the conceptualization and its ultimate consistency with the consequences associated with each cycle. The model contains these steps: (1) domain selection and definition; (2) purpose identification; (3) design; (4) administration; (5) interpretation; (6) evaluation; (7) analysis; (8) refocus; (9) planning/decision making; and (10) communication. In addition to these steps and the evaluation components associated with each, another dimension must be used to describe the assessment-evaluation cycle, the level of formality. At each step of the cycle, there is a separate set of evaluation components to be applied to maintain the most accurate assessment results possible. A chart outlines the evaluation components in terms of questions that reflect the information needed in relation to each component. Three examples describe assessment-evaluation at different levels of formality. There is little that is novel with respect to the processes of assessment and evaluation in this model, but the model does allow teachers to reflect on their practices, administrators to develop good data, trainers to integrate assessment and evaluation, and researchers to develop investigations. (Contains 3 tables and 42 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 458 263

The Assessment - Evaluation Cycle Model for Instruction

Michael J. Wilson

TM033426

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
M. J. Wilson
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

BEST COPY AVAILABLE

The Assessment - Evaluation Cycle Model for Instruction

Introduction

Teachers depend heavily on assessment and evaluation. To help their students grow, teachers must be constantly involved in assessing and evaluating a variety of things with direct impact on students, instruction, and the classroom environment. To help themselves grow, teachers must also include assessments of their own development and understanding of educational concepts related to instruction and the way their students learn. Each element of the teaching-learning process is predicated on teachers knowing and understanding the relative effect of a myriad of influences on their students. Assessment and evaluation is the path through which teachers gain the basic knowledge and enhanced understanding needed to address their students needs, organize and improve their instruction, and enhance themselves professionally. The success of teachers' assessment-evaluation efforts depends, however, on the way teachers conceptualize the various elements of assessment-evaluation process and how these elements fit together within the context of instructional practice.

Reason for an Assessment-Evaluation Model

The components of the assessment-evaluation process have been described in numerous texts (Airasian, 1996; Busick & Stiggins, 1997; Gedler, 1999; Haladyna, 1997; Kiester, 1999; Linn & Grondlund, 2000; MacMillan, 1997; McTighe, 1998; Murphy, 1999; Popham, 1999; Stiggins, 1997, 1998; Torrance, 1998; Wheeler, 1993; Wortham, 1996; Worthen et.al., 1999, etc.) and countless articles (Ancess, 1994; Arter, 1999; Brookhart, 1995, 1999; Bussert-Webb, 2000; Dunbar, 1992; Glaser & Silver, 1994; Haney, 1996; Lynch & Davidson, 1994; McMillan, et.al., 1999; Marzano, 1998; Mehrens, 1991a, 1991b, 1998; Messick, 1996; Odafe, 1998; Plake, et.al., 1993; Schurr, 1998; Shephard, 1997; Sternberg, 1997, 1998; Taylor, 1990; Valencia, 1997; Terwilliger, 1997; Wiggins, 1991, etc.). The important elements in the assessment-evaluation process have been addressed in these publications and articles and in many others not listed above. These documents provide the different aspects of educational assessment without providing a connection between the elements of assessment or showing how they connect to instruction. For example, in Popham's "Classroom Assessment" there are good readable discussions of the concepts of reliability, validity, and bias and a variety of assessments elements such as blue prints, different kinds of assessment formats and scoring processes. The book does not, however, contain a sequential process through which a teacher might proceed in order to accomplish the assessment. The model proposed here attempts to provide such a sequence, one which shows a cohesive approach to the integrated instructional use of the elements found in the assessment-evaluation literature such as shown above.

Possible Model Uses

Such a model would be useful in a variety of ways including the development of research to support a better understanding of the assessment-evaluation process in instruction,

identifying ways to improve the process as a whole, and improving training to support increased teacher facility and use of the process.

Through the use of a model such as the one proposed in this paper, investigations could focus on the effect of different elements of the model, on the accuracy of instructional assessment, the influence of each element of the assessment-evaluation process on instructional practices, the level of teachers' understanding and use of the different elements of the model, and many other aspects of the assessment-evaluation process. With a model forming the basis for future research on elements of the assessment-evaluation process as well as associated elements of instruction, the education community could gradually improve the model and with it our understanding of the way or ways that assessment effectively addresses or supports important aspects of the teaching-learning process. Examining one step in the aspect of the model, such as analysis, comprehensive assessment-evaluation research could:

- Investigate whether and in what form analysis does in fact act as described
- Determine what happens to the assessment-evaluation process with or without analysis
- Describe how different aspects of analysis are related to other elements of the assessment-evaluation process.
- Examine causative relationships between analysis and instruction.

Without a model, theory related to instructional aspects of the assessment-evaluation processes has been forced to address the separate elements or pieces of the process instead of addressing them in an integrated fashion. For some time now, investigators have examined how effectively teachers use elements like performance assessment in the classroom without connecting their findings to any comprehensive model of assessment-evaluation. As in the study of other human processes, the use of models forms the basis for advances through careful study and eventual understanding of the process.

For the lack of a model, transferable, trainable knowledge about the integrated aspects of the different steps in the assessment-evaluation process has not existed such that teachers do not have any clearly defined, research-based sequence through which to proceed when conducting assessment. To address teachers' lack of understanding of the assessment-evaluation process, a model would support many aspects of teacher training in this area. When devising training, for example, a model would give trainers a basis on which to relate the various elements of the process. As the model improves, the training could be more directive and provide a greater degree of clarity than currently exists. Such a model would allow investigators to determine whether training in particular aspects of the model provided the desired results. Teacher's assessment-evaluation practices could be examined and through different aspects of training programs, such as coaching, teachers could be provided with feedback that would help them improve their assessment-evaluation processes. Finally, a model could be the basis for training in college teacher development programs before teachers ever reach the classroom.

In the model proposed here, assessment-evaluation as practiced in instruction is described as a cycle. The term cycle suggests the completion of a sequence of recurring successive

events. This term is used because without following a particular set of steps, the results of the process become less valid and therefore less useful to instruction. In the model, these steps must be clearly defined with the interconnections delineated. The assessment-evaluation steps described in this article are designed to support an acceptable level of quality in an instructional assessment-evaluation process. Messick (1989), in his chapter on validity, indicated that all aspects of validity have as their base the construct underlying the assessment. The cyclic model described here begins with a thorough treatment of the domain and purpose, fundamental elements of assessment related construct. Messick also supports the importance of assessment-evaluation consequences. This model addresses consequences as an important culminating step in each assessment-evaluation cycle. The success of each cycle, then, depends on the conceptualization and its ultimate consistency with the consequences associated with the cycle.

Definitions

In the developmental of models, the importance of definitions cannot be overstated. A discussion in which definitions are not clear usually results in conclusions which must be held suspect. This proposed model of an assessment-evaluation cycle, requires somewhat different interpretations of terms relative to their use in other assessment literature. To avoid confusion and provide a clearer understanding for the following discussion, definitions of key terms are provided below.

- Assessment – Extracting information about student's cognition, affect and conation from what they say and do. We can assess by observing, by questioning, by listening, by measuring, by testing, etc.

- Evaluation – An evaluation is the process of judging. It includes selecting and defining that which is to be judged, and a standard or standards against which to make the judgment.

- Measurement – The aspects of assessment in which students are placed on some sort of scale or continuum that provides relative results such that all students involved in an assessment instrument can be placed somewhere on the continuum relevant to the constructs underlying the assessment. The scale can be as simple as mastered/not mastered or as complex as a multi-dimensional IRT calibration.

- Testing – A method of more formal measurement involving the use of an instrument or protocol and requiring components that provide increased technical efficiency such as standardization or scaling. For example, a quiz that has twenty multiple-choice questions on a specific subject would be a test. Assessing a student's writing ability by examining a homework paper or assessing the success of a lesson in the classroom through the quick use of some carefully selected question would not be termed a test.

Overview of the Assessment-Evaluation Cycle Model

The following are the proposed steps in the educational assessment-evaluation cycle as described in this article. As the title suggests, each step of the model has a relevant evaluation component. The steps will be described first, then the evaluation components of the cycle will be delineated with its corresponding step.

- Domain selection and definition
- Purpose identification
- Design
- Administration
- Interpretation
- Evaluation
- Analysis
- Refocus
- Planning/Decision making
- Communication

Formality

In addition to steps and evaluation components discussed above, there is another dimension that must be used to describe the assessment-evaluation cycle as it exists in the instructional environment – the level of formality. Formality is defined with respect to the level of time, rigor and technical considerations imposed on the development, administration, interpretation, and use of an assessment. In a low-formal assessment, one pass through the cycle may take place in matter of minutes. In a medium-formal cycle, the cycle may take several hours. In a relatively high-formal cycle, it may take months or years. The more time, effort, and technology that is invested in the cycle, the more formal the assessment-evaluation becomes. So an assessment developed in a few seconds, requiring a minute or two to move from purpose/domain to the ultimate use and communication of the interpretations and analysis would have a very low formality level. A norm referenced achievement test, on the other hand, which may take several years to develop and many hours to score, evaluate, analyze, plan, and communicate would represent an assessment that is high on the formality continuum. Generally, tests tend to be mid- to high-formal assessments while assessments that do not attempt to measure some aspect of a domain are usually low- to mid-formal.

Examples of level of formality

The following are some examples of different levels of formality in the assessment cycle. These examples have been arranged from low-formal to high-formal.

Low-Formal

- A teacher asking students questions about cells to determine the group's general level of knowledge of cells before launching into a lesson about different kinds of human cells.

- A teacher observing students reading aloud in the context of a group of students to determine whether they are involved in the story.

- The teacher listening for inaccuracies in pronunciation or intonation.

Mid-Formal

- A teacher doing a formal observation of a speech using specific trait rubrics that have been explained and demonstrated as the students prepared the speeches.

- A language quiz on the use of verbs in different tenses given at the end of a unit to determine how well the students can use the information gained to improve their written language and spoken language.

- A physical education final in which students are asked to demonstrate their skill and all they have learned about playing the game of tennis.

High-Formal

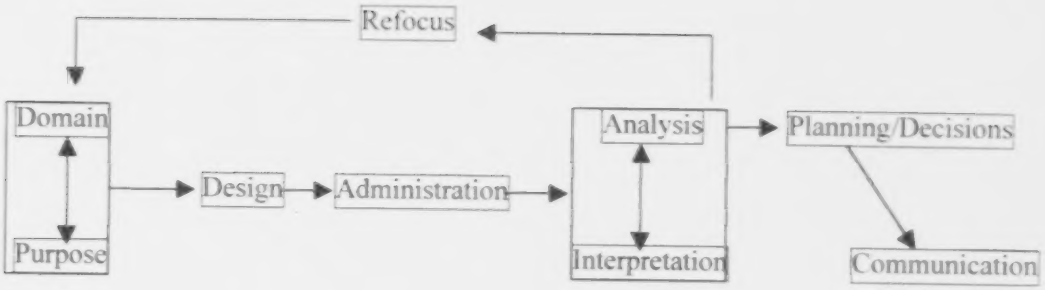
- A district assessment that provides teachers with performance standards that are stated in the district curriculum. Each student takes a test as the teacher reads a from a script of instructions so that the test is standardized and consistent for all test-takers.

- A commercial norm-referenced achievement test is administered to students. Again, there is a script to ensure a standard administration among all students. This test is composed of items that have been piloted and selected then administered to a large norming population of students across the nation.

- An authentic assessment given across a district to determine whether debate students are performing to their optimal potential. A set of scoring guides for all teachers is carefully produced and those who score on the test are given training and their scoring is recalibrated at regular intervals.

Model Elements

On the following page is a graphic depiction of the assessment-evaluation cycle that is outlined on page four. The boxes contain each of the steps of the cycle and the arrows show the flow from one element to the other. In some instances, arrows go in two ways. These arrows are meant to show that the two elements may be occurring simultaneously and that one element is influencing the function of the other. In other parts of the model the arrows proceed from one step in the cycle to the next. These one-way arrows are to show that the step or pair of steps is dependent on the preceding step. So according to the assessment cycle model shown on page six, each assessment cycle begins with a simultaneous and interactive consideration of the domain and the purpose of the assessment. After the domain and purpose have been decided upon, the next step is the design of the assessment followed by the administration and a simultaneous, interactive interpretation and analysis of the assessment results. After the interpretation and analysis steps, the assessment is used in planning and decision making and finally, the plan or decision is communicated to the appropriate person or persons.



Overview of Assessment Cycle

Below are explanations for each element of the assessment cycle as depicted above in the order in which they occur in the model.

Domain

A domain is a description of mental activity or processes, i.e., something that cannot be observed because it is taking place in the mind, but the results of which can be observed in what students do and say. For example, reading is primarily a mental activity, which can only be observed through activities like reading aloud, answering questions about what has been read or the student identifying the strategy, or strategies he or she uses to comprehend written material. Most assessment in education focuses on mental activities and therefore must be based on descriptions of mental domains. These descriptions are to a great extent hypothetical but the better they approximate what is actually occurring mentally, the more useful they become as the basis for the assessment cycle.

Domains can be relatively large such as the domain entitled science, but they can also be much smaller when they are thought of as portions of the larger domains. In the following discussions, these portions of larger domains will be referred to as sub-domains. The size and content of the sub-domain depends on the definition. For example the domain of social science could include a sub-domain of history which is still relatively large and can be diminished in size by focusing on specific aspects of history such as a specific year or element of history.

Domains can contain a variety of dimensions. In addition to the definition of the actual content in domains and sub-domains as described above, it is possible to include the dimensions of cognition, affect, level of development, type student response. So for example, in the domain of history we might want to assess a students problem solving ability (cognitive dimension) and commitment (affective dimension) with reference to the sub-domain of constitutional politics (sub-domain) using a constructed response (student response dimension).

Purpose

Purpose is the reason for designing and conducting an assessment, knowing why you are assessing whatever it is you are assessing. To develop a purpose for an assessment is to plan in advance how and in what context the assessment will be used. Being without a clear understanding of the purpose means the teacher is collecting information without identifying the reason for a particular assessment-evaluation cycle. For example, a teacher might give students an essay test with an expository writing topic not knowing what the test is to be used for. Then decide after collecting student responses to make decisions relevant to a narrative writing, the expository writing assessment-evaluation would be relatively worthless. Deciding on a purpose first with reference to the decisions, planning, and communications provides the teacher a much clearer path through cycle steps of design, administration and interpretation and the ultimate use of assessment results. To paraphrase the Cheshire Cat in Alice in Wonderland, if you do not know your purpose with an assessment, any assessment will do. It also follows that if the purpose is decided upon after the design and administration of the instrument, the assessment information is less likely to support the purpose.

There are a number of possible purposes. The chart below provides a conceptualization of the relation between the different levels of focus of assessment purpose relative to their location with respect to the instructional process. Each letter on the chart represents a different aspect of assessment purpose. The three columns on the right represent the parts of instruction – before, during and after. Each of the three rows represents a different focus of assessment purpose. So for example letter A is represents assessment purposes related to before instruction with a student focus. Letter E represents purposes associated with during instruction that focus on program or instruction and letter I represents assessment purposes associated with after instruction focusing on the teacher.

ASSESSMENT-EVALUATION PURPOSES

Level of Focus	Purposes Associated with Location in Instruction		
	Before Instruction	During Instruction	After Instruction
Student	A	B	C
Program/Instruction	D	E	F
Teacher	G	H	I

Each of the nine letters in the chart above are provided with examples of the kinds of purposes associated with each one.

Purposes	Example Purposes in Relation to the Instruction Process
A	To determine whether the student is prepared To understand whether the student has prerequisite knowledge To communicate to the student what is expected by the end of the units or the end of the course As baseline against which to compare progress To probe in depth the problems a small group of students has with a

	particular part of the domain.
B	To determine whether the student is learning what is being taught in a part of a lesson, a lesson, or an entire unit To determine whether any students are doing better than others To determine whether student strategies used in instruction are effective
C	To establish which aspects of a domain were learned better than others To determine whether students have achieved objectives To give students a grade To determine whether the student is prepared to go on or must return to review certain aspects of what has just been taught
D	To determine whether the instruction is appropriate for a given group of students To understand the subject domain with respect to the subject for the teacher and the students
E	To determine whether the instruction is working as expected for all the students To determine whether the teaching strategies are being effective
F	To determine whether the instructional design was effective To determine whether the curriculum was effective To determine whether domain was accurately specified
G	To determine whether there are processes or knowledge the teacher must acquire to work with the students To determine whether there are processes or knowledge the teacher must acquire to work with the curriculum To determine what kinds of training the teacher might need
H	To determine whether instructional strategies are working To determine what kinds of training the teacher might need
I	To determine to what extent instructional strategies worked To determine what kinds of training the teacher might need the next time this piece of instruction is used

Selection/Development of a Design

An assessment design is something that elicits observable student behavior. For the mental domain to be observed, students must do or say something in response to activities designed with reference to a particular domain and for a particular purpose. For example, if the domain is grammar and the purpose is to determine what a student needs to know to write well, the teacher must observe the results of mental processes associated with grammar. The assessment design therefore must include student behaviors which reflect those aspects of grammar the teacher's domain of grammar suggests are associated with good writing.

An important aspect of addressing the student response needed to make a connection with the domain specified in the first step of the cycle is to select an assessment format. Different aspects of the domain are better assessed by particular kinds of formats. These

formats have described thoroughly in the literature cited in the first pages of this document and will not be addressed in detail here. These assessment formats include:

- True-False,
- Multiple-Choice,
- Open-Ended Response,
- Essay Prompts,
- Performance Assessment,
- Portfolio Assessment.
- Observation
- Rating
- Anecdotal reporting
- Etc.

In addition to format the assessment content must coincide with the content specified in the domain. Many more formal assessments go to great lengths to create a specification or blue print for the assessment that carefully delineates the sub-domain and assessment dimensions as well as providing examples of assessment questions or items to be used as examples as the assessment is developed. In less formal assessment, there is not enough time to be so detailed in the planning of the design. To the extent that such planning can be made available through curricula and other planning materials, the accuracy of the assessment through the understanding and specification of the domain is enhanced and expedited.

Level of standardization of the assessment design is another important concern at the design step that must be followed with standardization in the administration step as well. Standardization is the process of ensuring that each student's response is to the same set of stimuli. The responses to assessments that have been standardized are more easily compared to one another those assessments which do not have standardized elements.

If a format is correctly chosen and the domain assessment content match is accurate, a teacher should be able to obtain the domain information they specified in the purpose by either listening to or observing student's responses. Teachers that have the most thorough knowledge and understanding of a domain will be able to either select or develop the kinds of questions they need to assess a particular domain. Accurate assessment of the domain depends on teachers understanding both the components of the domain and how they related to the kinds of student behaviors available through the selected format.

Administration of Assessment

The administration is dependent on the outcome of the assessment design step. The selection of the format directly effects the kinds of behaviors that are to be observed in the administration of the design selected or developed. If for example, a multiple-choice format is selected, the teacher merely has to observe that students respond to the questions without assistance from their fellow students or some other kind of assistance. If, on the other hand, the teacher has selected a performance assessment format, it may be necessary to interact with the students in order that they might obtain necessary materials,

to give specific instructions at different parts of the process, or maintain specific guidelines during the process.

The administration of an assessment is the seemingly simple act of having students respond to the assessment as designed. If a student does not respond to the assessment, then the teacher can gain no understanding of the underlying mental process and will therefore be unable to understand anything about the domain of interest. For teachers allowing students to respond to questions without providing hints or suggestions can be the most difficult part of the process, especially when this means letting students make mistakes or show limitations related to just applied instruction. There is always the urge to give a hint or assist the memory process in some way or other. It is particularly difficult to let students make mistakes in low-formal administrations because less standardization is involved and modifications of questions and teacher reactions to questions can easily slip into the assessment-evaluation process. A teacher who helps a student during an assessment of a given domain, however, may take the risk of developing an interpretation and analysis from students' responses that might not accurately reflect the students' actual mental processes related to the domain of interest.

The way that assessment typically overcomes the biases described above is standardization of the assessment instrument – everyone responds to the same set of questions - and the administration procedures – everyone responding gets the same set of instructions and other stimuli associated with the assessment. Standardization is important if one of the purposes includes the need to make comparisons between or among students. Student responses that result from different administration procedures may be different just because of the differences in what teachers say or do or difference in the environment in which the procedure was performed. Therefore, teachers who want to make these comparisons must observe the same behavior from one student to the next. Standardizing or ensuring consistency of response from one student to the next would therefore would be an important element of the administration portions of the cycle.

Interpretation

Interpretation includes several components but the purpose for this step is to take student responses generated during administration and imbue them with meaning. These steps include scoring, referencing, and recording. Not all assessments require all of the steps, but after student responses have been collected, these responses must proceed through the interpretation process before they are useful. In the low-formal process of observation or questions asking, interpretation may include nothing more than referencing to a concept or expectation or making a mental notes about the presence or absence of a given skill. Scoring will generally only occur in assessments that are developed in the form of a measure or a test.

Scoring

Scoring is the process of placing students on a domain related continuum. This continuum can be numerically described or categorical. A numerical scoring suggests a relative degree of accuracy and mathematical manipulation. Categorical scoring suggests

that the assessment results can be compared or ordered but they preclude any mathematical manipulation. Depending on whether they are numerical or categorical interpretation of student responses may include:

- Summing points from correct or partially correct responses.
- Ordering scores from lowest to highest
- Discarding extreme scores that do not seem to represent actual student performance.
- Developing a rubric score. This process is a combination of both the scoring and referencing process because with a rubric score the teacher would be simultaneously applying a level of performance that could be placed on a continuum and implying meaning by referencing to a judgment process. For example
- Weighting different scores depending on the level of use in instruction or level of importance in the domain.
- Transforming number correct scores to percent or percentile rank or to standard scores.

Referencing

Referencing is the process of comparing students' responses with something external to the response in order to better understand the response. The response may be compared to a norm groups' score continuum or to important attributes such as examples of successful achievement or good scholarship or score points established to represent a satisfactory level of achievement. Assessments based on measures or tests may use reference comparison on normative mathematical continua whereas assessments that are categorical not will need this kind of referencing. The following is a list of possible referencing contexts:

- Example – when a teacher assessment determines whether a particular student behavior is similar to another student performance
- Expectation – When a teacher assessment relates a student response to a generalized notion the teacher holds about the domain
- Prototype, - Relating a students response to a model that has been careful design to describe a particular process or developmental sequence.
- Criterion – A given level of correct response beyond or within which a certain grade is given to a performance
- Baseline, - An initial score level against which subsequent scores can be compared to determine whether there has been improvement in the performance
- Norm – The relative score relationships in a general population such that a particular student score on a given instrument can be related to the population's scores to get a sense of the level of the performance
- Standard – A particular point on a continuum of scores above which mastery or some other level of performance is attained
- Domain - A level that reflects to what proportion of the domain has been achieved.

Recording

When recording responses, the teacher needs to think about the metric – what kinds of numerical or categorical representations to use, format or form of the data, and how the recording will be used in the subsequent analysis in order to accomplish the purpose established at the beginning of the cycle. The following are some of the decisions that need to be made in the recording process.

- Metric – Are scores going to be recorded using numbers, graduated descriptions, rubrics, or some other more complex numeric form such as a percentile ranks, or standard scores.
- Electronic – Electronic recording devices could include spread sheets, grading packages, data bases, data warehouses.
- Format – The format of the data in the interpretation step should facilitate or at least not interfere with the next step, analysis. If for example, the data needs to be broken into groups and mean averages generated, then the data format should be organized to facilitate this analysis. If the purpose calls for group analysis, then the recording format should facilitate the analysis with group coding as well as score interpretation data. Mental or anecdotal notes format can be used with assessments that are not in the form of measures or tests by either being quickly placed in memory or formalized into written anecdotal notes that describe observations made of students' verbal or physical responses.

Analysis

After the interpretation has been completed, and to prepare for planning or decision-making and communicating, the interpretations of students' responses can be examined for patterns that might indicate student needs or provide program or teacher information relevant to the purpose of the assessment. Once the recorded data are set up and the different aspects of the analysis properly identified, it is possible to accomplish most analysis of subsets of students or sub-domains related to the subject being assessed. For example, if the purpose of the assessment is to determine what aspects of a given domain a teacher is particularly good at in his or her instruction, the analysis might include an examination of a breakdown of specific subject sub-domains across different levels of student performance. The analysis might also examine the patterns of interpretations from students' responses, among specific groups of students or across the entire class of students. In another example, the teacher might notice that all of the mathematics problems containing percent and decimal numbers were answered incorrectly by a particular student or a subset of students. Focusing on subject sub-domain, a teacher could look across the entire class and determine which aspects of mathematics the entire class showed the smallest number of correct responses. Analyses may vary widely depending on the purpose established at the beginning of the cycle.

To conduct an analysis, the teacher must depend on a good understanding of the domain and purpose underlying the assessment. For example, if the purpose is to determine whether students are ready to begin instruction at a certain level of reading and the instrument design focuses on aspects of the domain that are irrelevant to all but one or two aspects of reading readiness, the analysis of the assessment may be worth very little. In addition, the subsequent planning and communications accomplished from the

assessment will also be limited if the design has not addressed important aspects of the domain.

The teacher with a good knowledge of the assessment domain should be able to evaluate the assessment results and judge whether or not they provide the kind of information needed relevant to purpose. If a teacher can judge the extent to which the domain has been assessed, this information can be used to form an idea of how extensively the results of the analysis can be generalized to the domain that has been assessed. If certain aspects of the domain have not been well represented by the assessment, that brings us to the next step in the assessment-evaluation cycle, Refocusing.

Refocus

After an analysis has been conducted, it may become apparent that there is not enough information on one or more students to accomplish the original purpose, or the information available is contradictory and the lack of clarity would interfere with the purpose. When it becomes clear that more information is needed, it is time to return to the beginning, to the first step of the cycle. The new purpose, at this point, is to refine or clarify the assessment with respect to the results of the analysis or perhaps even the interpretation steps in the cycle. The new domain definition would be generated based upon that aspect of the subject that needed to be clarified or refined on the original pass through the cycle. After the new domain and purpose are established, the teacher continues on through the rest of the cycle steps until she comes again to the analysis step. At which time, she can determine whether or not the information resulting is sufficient to the purpose originally established in the initial cycle. For example, some students have incorrectly answered questions about the main idea and supporting details of a written report on whales. The teacher wants a better idea of what problems might exist that are hindering the students ability to understand a text on whales. In her analysis of the problem, she refocuses on the students ability to read the words and understand the vocabulary. After finding no problems in reading and understanding the vocabulary, she refocuses on assessing whether the students were familiar with the concepts in the story, such as the concept of whale, migration, ocean, etc. Finally, she might refocus on the strategies that the students used to arrive at the answers they provided in the assessment. All of these iterations of the refocusing cycle were used because of their connection to the domain of reading and the original purpose of the assessment. A refocus of the assessment may not be needed, but it should always be considered to ensure that plans and decisions based on analyses of assessment data provide accurate information.

Planning/Decision Making

Planning is the act of accomplishing assessment purposes through the use of the analyses of assessment interpretations. Planning/decision making may include the development, selection or refinement of instructional strategies, designing or selecting student learning strategies or student thinking strategies, consideration of teacher training, modifications of classroom environments, planning of parent interactions or materials acquisition, etc. In sum, planning is the act of using the assessment information gathered for a particular purpose and making either decisions or plans or both with respect to that purpose. Accuracy in all the steps of the assessment-evaluation cycle will provide decisions and

plans with a firmest foundation possible. To the extent that planning is based on effective use of the cycle and especially consistent with the original purpose on which the design and administration are based, the consequences embodied in the planning and decision phase of the cycle will imbue the entire cycle with more valid results as recommended by Messick (1989) and his concern for consequences with reference to valid assessment.

Communication

Communication is the process of transferring the assessment results and related planning/decision making to the people for whom the assessment information was collected. As in all other instances, communication depends on the original purpose of the assessment. Communication may include:

- Teacher with self
- Teacher to student
- Teacher to parent
- Teacher to administrator
- Teacher to community

Communication of plans or decisions based on assessment integration and analysis should conform to the purpose. The following are examples of how the purpose and the assessment and resulting communication all need to be matched.

Purpose	Assessment	Communication
TO SELF: To determine whether a unit on gravity was working as designed	Unit test	Identifies aspects of the curriculum that need to be modified or removed and makes a mental note of the kind of changes that need to be made.
TO STUDENT To determine whether the student is prepared to begin cursive writing	Student is asked to copy a series of straight and curved lines	Individual conferences with the students to indicate that they are ready or to explain how they will become ready. Students who appear to be ready get a happy face sticker on their paper.
TO PARENTS: To show parents the progress that has been made in writing skills	Portfolio of rubric scored writing pieces done by students over the semester.	The teacher in conferences with parents points out with each individual's writing samples what has improved in the student's writing
TO ADMINISTRATOR: To show the effect of a new approach to science instruction	Past results of previous science tests and unit tests of new science program	Using the old and new science program test results, and the science domain, the teacher points out what has been learned with the old

		program versus the new program.
TO COMMUNITY To identify the need for training in a particular area of language arts.	Commercial test including the sub-domain of concern	Using the low student results in the sub-domain of interest, the teacher argues before the board meeting to increase training to address that aspect of language arts.

Communication can be accomplished in a number of different ways. These ways could include any one of the following:

- One on one conversation
- Parent conferences
- Peer conferences
- Written reports
- Comments on student papers
- Grades
- In group conversations
- Class discussion
- Demonstrations
- Etc.

Evaluation Components of Assessment Cycle

At each step in the cycle described above, there is a separate set of evaluation components that are applied in order to maintain the most accurate assessment results possible. Below is a chart that outlines these evaluation components which are couched in terms of questions that reflect the kind of information needed relative to the component.

Cycle Steps	Description of Steps	Possible Evaluation Questions
Domain	The specification of content and structure of what is to be assessed.	Which aspects of a subject are relevant and should therefore be included in a domain?
Purpose	Determining the reason for an assessment and how the assessment results will be used.	What reasons are important? What are the most important purposes? How to focus the purpose to make the best use of time and energy? How to judge effectiveness.
Design/Selection	Designing or selecting the kinds of student behaviors needed to reflect the purpose and domain.	What are the limitations of the behaviors selected? Will they provide the teacher with the kind of information needed?
Administration	Establishing processes and approaches to obtain objective interpretable student responses.	Is the level of objectivity sufficient to give an accurate picture of what the student can

		in fact do? Are there any teacher biases that might interfere with accurate results? Are there any conditions that existed to cause students to perform better or worse than they should have given their level of understanding, knowledge or skill?
Interpretation	Translating student responses into meaningful elements.	Does the interpretation consistently and accurately portray the capability, or level of knowledge or skills of the student? What inferences are appropriate? What are the standards? Where is the student in relation to the standard? What are interpretation biases that affected assessment results either positively or negatively?
Analysis	Observing and distinguishing patterns.	Is the result what is expected given the purpose and the domain?
Refocus	Returning to beginning of cycle to obtain clarifying information.	Is there sufficient information in the analysis and interpretation to meet the purpose?
Planning/Decisions	Using the interpretation and analysis of student responses to plan and make decision about students, programs, instruction and teachers.	Is there sufficient assessment information to make the decision or for the plan to be successful?
Communications	Creating messages that communicate decisions and planning.	Is the assessment data in a form that makes sense? Is there enough explanation to ensure continued understanding of the material?

Examples

The following paragraphs contain three examples of complete assessment cycles at different levels of formality. In each example, each in the the assessment-evaluation cycle is depicted. The three examples show three levels of formality described earlier in this paper.

Example One (low formal):
A teacher is in the middle of a lesson on the constitution and wants to know whether her explanation of the freedom of speech was understood by her students. So she asks some questions: "Does the freedom of speech mean you can say anything you want to another person or a group of people? Is a paper permitted to print whatever it wants? Does the right to free speech prohibit us from telling someone to hurt another person?" The domain of the assessment is government, the questions are from a sub-domain termed Bill of Rights. The purpose or reason for the questions is to determine whether the past twenty minutes of instruction on the freedom of speech in the Bill of Rights has been successful. The design of the assessment is based upon the need to obtain quick student responses and the questions are selected from the aspects of the domain covered in instruction. To administer the assessment, the teacher decides to have each student close his or her eyes and raise one, two, three, or four fingers in response to the multiple-choice type questions. One finger indicates response-choice A, two fingers indicate B, three fingers indicate C and four indicate not sure. Her interpretation of the responses will tell her which students appeared to be understanding the information included in the instruction and which do not. Part of her interpretation is a mental recording of which students did not understand the instruction from their responses to the three questions. She is able to determine from her analysis of the responses whether particular students understood and perhaps what aspects of the discussion need to be reviewed during the rest of the instructional time or during subsequent instructional periods.

By noticing that five students who had missed school the previous week did not correctly respond, her analysis suggests that these students are missing the information from the previous week that would be necessary to understand the current weeks lesson. She makes a decision to take the students aside and work with them on the concepts they need. She also decides not to refocus the question at this point due to the time factor – needing to finish the lesson before the end of the period. Later, she will obtain a better idea of the problems these five students are having when she refocuses this original assessment using questions on the relevant parts of the domain in small group instruction.

By looking around the room at the finger responses to her questions, the teacher interprets and analyzes the responses in reference to her lesson. During this process, she also notices that one question was answered incorrectly by a majority of students. Because she is familiar with the domain and instructional strategies that are relevant to that particular domain, she immediately adjusts her plans for the next half hour of her instruction on the Constitution. She quickly communicates the results of the assessment by noting that most students seemed to understand the first two concepts and explains what weaknesses she perceived through their responses and explains her decision to review a particular part of the Bill of Rights they just covered.

Example Two (medium-formal):

The teacher has decided that the purpose for his assessment of student writing is to establish a series of baselines from which both he and the students can determine whether instruction has been successful. The teacher has carefully constructed an outline of the dimensions of the writing domain and has decided that he will address the part of the domain he calls persuasive/argumentative writing. He carefully constructs five writing tasks that include as many different aspects of this sub-domain as possible. Each of these tasks relates to one or more of the different writing strategies that he will incorporate into his instruction. Each student's response will be holistically scored so that the interpretation process will provide a simultaneous application of score and referencing to a level of mastery. In the scoring, a 0 equals no mastery, 1 equals partial mastery, 2 equals mastery and 3 equals advanced mastery. Each score point is accompanied by a description and a set of example papers the teacher refers to as anchor papers. Each task will be administered during a forty-five minute first draft writing session. He will ask the students to take two of the writings and go through several drafts until each student feels she or he has a finished piece of writing from the two original drafts. The scores for all five papers will be recorded in a spread sheet for the teacher to review. All the drafts will be kept in individual student files. The teacher will analyze the scored writing samples to determine which students and which aspects of argumentative/persuasive writing need additional instructional help. After the analysis, he decides that several students may need considerable assistance and refocuses the assessment by obtaining a few more writing samples on these students so that he can confirm the problems that he has observed and also obtains a more comprehensive picture of the overall writing process of these students. At several points during the unit on persuasive writing he will ask his students to develop analyses of their own writing by identifying the progression of their arguments and strategies for persuading their readers. Listening to students' analyses of their own work, he will be able to determine whether students have successfully understood much of the instruction. After the analysis of the writing tasks, the teacher will communicate to each student the instructional plans he feels are needed to strengthen group's writing skills in argumentative/persuasive writing.

Example Three (high-formal)

A teacher is administering a fourth grade state test. The examination questions and publications require about one year to produce using committees of teachers and professional item writers. The purpose for the administration of the test is two fold. First it lets students, teachers, parents and school administrators know whether or not students are achieving state standards at the level expected by the state. Second, it provides a tool with which to make the school accountable for the achievement of the academic standards incorporated into the test developed by the state. The examination is administered over a four day period. The state has prepared detailed specifications that outline the dimensions of the domains included in the different sections of the test. One domain is mathematics and another is language arts which includes the sub-domains of reading, writing, speaking, and listening. The interpretive aspects of the test reports show the students' level of competency in language arts and writing but do not provide detailed sub-domain information. These reports allow teachers and the school to analyze broad aspects of language arts that need improvement across the fourth grade population. To

develop a more specific refocus of the different aspects of the sub-domains associated with language arts, however, the district must develop instruments that allow a more detailed analysis of the subject domains assessed by the state test. The state scoring provides an interpretation of the students' scores in the form of competency levels that suggest whether each student's performance is either satisfactory or unsatisfactory. Using the data from the state interpretations, the analyses and refocusing using the district developed assessment, the schools with the help of the district personnel are able to identify aspects of their mathematics program that need additional planning. After the planning is accomplished, the committees involved assist the superintendent to communicate the modifications that need to be made to the curriculum and instructional procedures.

Conclusion

The approach suggested in the above model provides very little that is novel with respect to the processes of assessment and evaluation. What is provided is a simple model that teachers can use to reflect on the various aspects of their assessment-evaluation practices, administrators can use to develop good data, trainers can use to integrate the various aspects of assessment-evaluation, and researchers can use to develop investigations that will strengthen this entire area.

REFERENCES

Airasian, P. (1996) *Assessment in the Classroom.* New York, McGraw Hill.

Ancess, (1994) *Authentic Teaching, Learning, and Assessment with New English Learners at International High School. A Series on Authentic Assessment and Accountability*, NCREST, Teachers College, Columbia University, New York.

Arter, (1999) Teaching about Performance Assessment, *Educational Measurement: Issues and Practices*, Summer, 18, 20-44.

Brookhart, S. (1995) Effects of the classroom assessment environment on achievement in mathematics and science. *ERIC* (ED382687)

Brookhart, S. (1999) Testing about the role of classroom assessment in student motivation and achievement. *Applied Measurement in Education*, 12, 409-425.

Busick, K. & Stiggins, R. (1997) *Making Connections: Case Studies for "Student-Centered Classroom Assessments.* Portland, OR, Assessment Training Institute.

Bussert-Webb, K. (2000) Did my holistic teaching help students' standardized test scores? Journal of Adolescent & Adult Literacy. 43(6), 572-573.

Dunbar, S. (1992) Classroom assessment. *Journal of Educational Measurement*. 29(4), 355-357.

Gedler, M. (1999) *Classroom Assessment and Learning.* New York, Addison Wesley Longman.

Glaser, R. & Silver, E. (1994) Assessment, testing, and instruction: Retrospect and prospect. *ERIC* (ED381549)

Haladyna, T. (1997) *Writing Test Items to Evaluate Higher Order Thinking.* Needham Heights, MA, Allyn & Bacon Pub. Co.

Haney, W. (1996) Focusing assessments on teaching and learning. *New Schools, New Communities*, 12(2), 11-20.

Kiester, J. (1999) *Blowing Away the State Writing Assessment Test. Four Steps to Better Scores for Teachers of All Levels.* Gainesville, FL, Maupin House Publishing, Inc.

Linn, R. & Grondlund, N. (2000) *Measurement and Assessment in Teaching.* Des Moines, IA, Prentice-Hall

Lynch, B. & Davidson, F. (1994) Criterion-referenced language test development: Linking curricula, teachers, and tests. *TESOL Quarterly*, 28(4), 727-743.

MacMillan, J. (1997) *Classroom Assessment. Principles and Practices for Effective Instruction.* Needham Heights, MA, Allyn & Bacon.

Marzano, R. (1998) Rethinking tests and performance tasks. *School Administrator* 11(55), 10-12.

McMillan, J., Myran, S., Workman, D. (1999) The impact of mandated statewide testing on teachers' classroom assessment and instructional practices. *ERIC* (ED431041).

McTighe, J & Ferrara, S. (1998) *Assessing learning in the Classroom. Student Assessment Series.* Washington DC, National Education Association.

Mehrens, R, Popham, J, & Ryan, J. (1998) How to prepare students for performance assessments. *Educational Measurement: Issues and Practice* 17(1), 18-22.

Mehrens, R. (1991a) Defensible/indefensible instructional preparation for high stakes achievement tests: An exploratory trialogue *ERIC* (ED334202).

Mehrens, R. (1991b) Facts about samples, fantasies about domains. Educational *Measurement: Issues and Practices* 10(2), 23-25.

Messick, S. (1996) Validity and washback in language testing. ERIC (ED403277).

Messick, S (1989) Validity. In *Educational Measurement*. Linn, R. (ed.) New York, Addision Wesley Pub. Co.

Murphy, P. (1999) *Learners, Learning & Assessment. Learning, Curriculum and Assessment Series.* Thousand Oaks, CA, Sage Pub.

Odafe, V. (1998) Students generating test items: A teaching and assessment strategy. *Mathematics Teacher* 91(3), 198-202.

Plake, B. (1993) Assessment competencies of teachers: A national survey. Educational *Measurement: Issues and Practices* 12(4), 10-12,39.

Popham, J. (1999) *Classroom Assessment: What Teachers Need to Know.* Needham Heights, MA, Allyn & Bacon

Schurr, S. (1998) Teaching, enlightening: A guide to student assessment. *Schools in the Middle* 7(3), 22-27,30-31.

Shephard, L. (1997) Measuring achievement: What does it mean to test for robust understanding *ERIC* (ED415238).

Sternberg, R. (1997) What does it mean to be smart. *Educational Leadership* 54(6), 20-24.

Sternberg, R. (1998) Teaching and assessing for successful intelligence. *School Administrator* 55(1), 26-27, 30-31.

Stiggins, R. (1997) *Classroom Assessment for Student Success.* Washington DC, National Education Association.

Stiggins, R. (1998) *Student-Centered Classroom Assessment.* Old Tappan, NJ, Prentice Hall, Inc.

Taylor, D. (1990) Teaching without testing: Assessing the complexity of children's literacy learning. *English Education* 22(1), 4-74.

Terwilliger, J. (1997) Portfolios and classroom assessment: Some claims and questions. *ERIC* (ED419000)

Torrance, H. & Pryor, J. (1998) *Investigating Formative Assessment: Teaching, Learning and Assessment in the Classroom.* Florence, KY, Taylor & Francis, Inc.

Valencia, S. (1997) Authentic classroom assessment of early reading: Alternatives to standardized tests. *Preventing School Failure* 41(2), 63-70.

Wheeler, P. (1993) *Methods for Assessing Performance.* Livermore, CA, EREAPA Associates.

Wiggins, G. (1991) Toward one system of education: Assessing to improve, not merely audit *ERIC* (ED348400).

Wortham, S. (1996) *The Integrated Classroom: The Assessment-Curriculum Link in Early Childhood Education.* Upper Saddle River, NJ, Prentice Hall Business Publishing.

Worthen B. (1999) *Measurement and Evaluation in the Schools.* New York, Addison Wesley Pub. Co.